

# 基于区间统计分析的迭代式信息采集模型的构建研究

张梦川 张濛升 戴冉 马明

郑州市烟草公司卷烟配送中心 河南 郑州 450000

**【摘要】**：为提升卷烟价格信息采集的科学性与稳定性，本文基于国家局选点机制构建一套基于区间统计分析的迭代修正模型，模型借助客户分类模拟与精度动态调整策略，逐步修正商户反馈的价格判断偏差。在理论层面引入大数定律与非合作博弈等支撑模型构建，在实验层面依靠三轮模型迭代完成从 V1.0 到 V3.0 的优化。实际应用中模型在 754 户样本基础上进行六轮数据采集与修正，验证其中低价位段的高容错性与强收敛性，研究结果显示该模型能有效提升价格数据采集的准确性，为烟草行业数字化管理提供可靠的数据支撑。

**【关键词】**：价格修正模型；问卷采集；区间统计；迭代优化；烟草行业

DOI:10.12417/2705-1358.25.18.065

## 1 引言

### 1.1 研究背景与意义

在卷烟价格信息采集过程中，不同渠道数据如国家局直与白名单客户反馈存在采集周期不一致和数据质量波动等问题，难以满足当前政策调控与市场监管对精准数据的需求。本研究以国家局选点样本为基础，采用多轮问卷交互机制，对卷烟价格数据进行动态纠偏与优化，为烟草行业数字化转型与管理决策提供数据支撑。

### 1.2 研究现状

当前卷烟价格数据的研究虽具备一定的基础，但尚未建立统一且科学的价格数据体系。多元数据来源在采集方式和质量标准等方面存在明显差异，导致数据整合面临结构不一和标准不统一的问题，价格数据一旦出现偏差，缺乏有效的修正工具和算法支撑，导致数据质量长期得不到有效保障。

### 1.3 研究预期成果与创新点

本课题致力于构建基于数学模型的价格数据采集修正体系，以提升营销子系统采集数据的科学性与稳定性，借助低代码平台与多轮自动交互机制，实现数据从采集到处理、从修正到反馈的全流程闭环。在模型设计上引入大数定律与不可能三角理论，对客户行为感知和市场博弈特征进行建模，为修正逻辑提供理论支持。

## 2 理论基础和技术路线

### 2.1 理论基础

(1) 大数定律。大数定律是统计学中描述频率稳定性的

基本原理，体现了随机变量在大量重复实验中呈现出的稳定趋势。当试验次数足够多时，随机事件的频率趋近于理论概率，反映出个体偶然性在总体中逐渐被平均化的规律。

(2) 商品供求规律。商品供求规律揭示了供给、需求与价格之间的互动关系，在经济活动中形成价格波动的核心机制。价格变动反过来也会影响供求行为，高价激发供给与抑制需求，低价则抑制供给以及推动消费，在卷烟市场中同一品规在不同区域存在价格差异。

(3) 商品价格构成。价格构成是指组成产品价格的各个要素及其在价格中的组成情况。从市场营销角度来看，产品价格的具体构成为：

$$\text{价格} = \text{生产成本} + \text{流通费用} + \text{税金} + \text{利润} \quad (2-1)$$

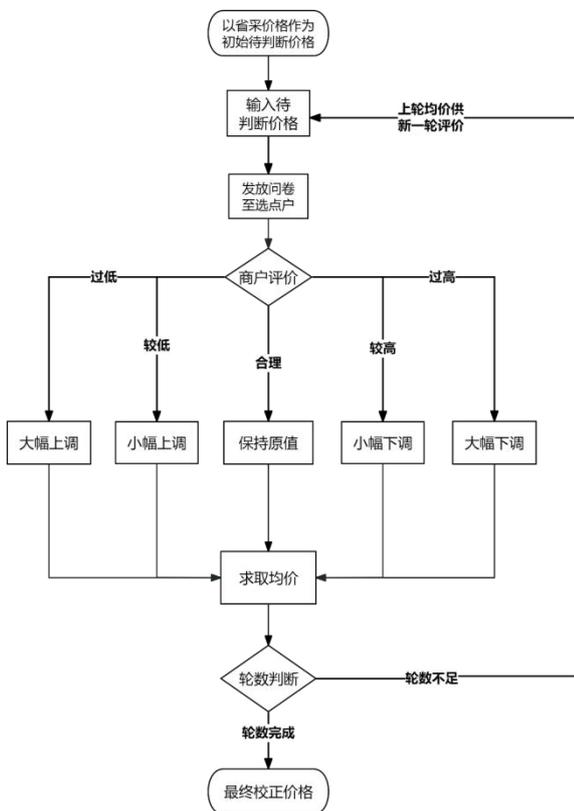
价格构成包括生产成本、流通费用、税金与利润四部分。生产成本体现为物资消耗与劳动报酬的货币表现，流通费用涉及商品在流通领域的耗用与时间空间关系，税金作为国家调控手段影响价格水平，利润则反映企业创造的价值，是再生产的重要资金来源。

(4) 非合作博弈。分支，关注各参与方在资源或利益存在冲突的情境中如何独立作出策略选择以实现自身收益最大化。在这一博弈框架下个体不寻求联盟或协议，而是基于自身判断和局部信息独立决策。

### 2.2 技术路线

依据区间统计分析和迭代的思想，我们设计了多轮交互修正的技术路线，如图 2-1 所示：

图 2-1 技术路线图



### 3 假设论证

本课题结合理论构建与模拟实验设计，在实践调研基础上提出修正模型，并对其有效性展开系统验证。研究依据前述理论框架设定核心前提，明确建模边界条件。

#### 3.1 基于选点原则提出的假设

基于对国家局权威性、全面性的信任，同时考虑到对于既往选点路径的延续的基础之上工作展开的便利性，保证采集选点的科学性与可靠性，本研究选点将采取国家局选点标准，故有以下 1、2 前设：

假设一：国家局选点标准是科学的，样本能够反映市场整体。国家局的选点标准建立在系统调研与数据分析基础上，涵盖地理分布与商户类型等多维因素，具有高度的代表性与规范性，标准制定过程结合专家意见和市场实际，保证所选商户样本能较为全面地覆盖行业特征。假设二：国家局选点标准下不同的样本对整体的反映是近乎一致的。国家局的选点标准建立在系统调研与数据分析基础上，涵盖地理分布、经济结构、商户类型等多维因素，具有高度的代表性与规范性。标准制定过程结合专家意见和市场实际，保证确保所选商户样本能较为全面地覆盖行业特征。

这两个假设的成立是本课题后续研究的基础，对河南省烟草局采集的重点品规卷烟价格进行准确地修正和分析。

#### 3.2 基于产品流通提出的假设

从价格采集理论支撑角度入手，卷烟作为一种商品，其价格是否受市场供求关系影响是我们首要考虑的问题，所以提出假设三：

假设三：卷烟作为市场流通的产品，符合经济学市场规律，价格受市场供求关系影响。卷烟虽处于计划生产体系，但在终端市场上仍表现出典型的供求调节特征。非订满品规可由商户增订补充供应，订满品规因额度限制导致价格上升，表现出供不应求下的价格机制。例如“十渠”在年底缩减投放后出现价格上扬，促使商户释放库存。季节变动与库存水平也影响价格走势，淡季则加快出货提升销售回报，市场供给增加，价格趋稳。

### 4 模型构建与验证

#### 4.1 模型准备

##### 4.1.1 环境设置与客户模拟

本模型中实验环境的搭建旨在再现真实市场中客户对价格的感知与反馈过程，为反映市场异质性，将客户划分为三类：A 类（价格敏感型）、B 类（价格迟钝型）与 C 类（消极合作型）。A 类客户能够精准识别接近均衡价的市场信号，B 类客户的判断存在感知偏差，C 类客户反馈行为呈现明显随机性。模型中引入“上帝价格”作为设定的理论均衡价，在每一轮模拟中输入“判断价格”，由各类客户基于感知判断区间反馈修正值，并据此完成价格修正的迭代过程。客户的价格感知模拟遵循：A 类围绕上帝价格±小幅范围波动，B 类在较大合理范围内波动，C 类反馈结果则完全随机，模拟其非理性行为。

##### 4.1.2 参数设定与流程定义

###### (1) 模型验证与优化

①模型 V1.0 与 V2.0 验证效果分析。V1.0 模型在全量实验场景下（卷烟规格、误差容忍度、初始偏差组合与客户结构多元设置），共模拟 126 万次。实验发现，模型在高价位段表现较好，多数情况可在三步内修正完成；但中低价位段误差按比例取值的设定使精度要求过于苛刻，导致修正步数频繁达上限，影响模型可靠性。为解决这一问题，V2.0 模型引入“固定比例+固定值”的精度判定方式，使误差范围更加贴合实际业务需求。经 378 万次模拟验证，新模型在低价位段的修正成功率显著提高。例如在 GP=110、JC=0.01、固定值为 1 的条件下，四步内完成修正的比值提升至 97%以上，说明精度设定策略的

调整对模型性能有显著提升作用。

②模型 V3.0 优化策略与验证结果。考虑到 V2.0 在极端案例下依然存在修正能力边界问题，V3.0 模型进一步引入动态区间宽度调整机制。即依据价格区间自动调整判断区间范围，提升模型对极端高低价位卷烟修正能力。在 112.5 万次模拟实验中，V3.0 模型在各价位段均实现高可靠修正。以 GP=110 为例，不同初始判断值下的 LUj 最大修正步数均未超过 10，且绝大多数情境中可在 3~5 步内完成修正，模型稳定性显著提升。

## 5 模型应用与效果分析

本课题构建的 V3.0 模型基于区间统计与迭代算法，在实际应用中设计了一套以问卷调查为基础的数据采集与修正流程。选取 754 户商户作为数据来源，持续收集其对重点卷烟品规的价格判断数据，每轮反馈均进行精度筛选与修正处理。修正结果输入下一轮问卷，形成连续五日的动态数据更新链条。每一轮结果作为后续问卷基准，迭代输出修正数据，第六轮问卷作为模型准确性的验证工具，用以比对前五轮修正趋势与结果稳定性。若验证问卷反馈与五轮修正数据趋同，说明 V3.0 模型具备良好的价格判断校正能力。

### 5.1 问卷前期准备

#### 5.1.1 问卷需求

(1) 问卷检验需求。问卷需设置有效筛查机制，结合嵌入式识别项与数据统计方法，识别异常填答与逻辑矛盾，借助自评与复现题进行预设检验，结合马氏距离与 IP 记录等维度进行事后数据剖析，剔除无效样本。(2) 问卷投放需求。问卷结构应紧贴研究范围并兼顾商户填写体验，界面简洁易懂并减轻理解负担。客户经理提前了解问题设定并做好引导，首次问卷采用高频次回收，有助于压缩时间变量影响，增强数据修正模型的响应准确性。

#### 5.1.2 问卷设计

(1) 问卷检验设计。为减轻客户经理理解负担与商户作答压力，问卷检验环节未采用复杂设计。在保证 99%准确率的

前提下，模型可容纳不超过 50%的随机误差，具备较强容错性，适应实际应用需求。(2) 问卷有效性题目设计。问卷内置两项检验内容，一项为复现性测试，采用相同题目检测一致性，一项为反常设问，利用极端设定引导唯一合理选项，用以识别随机或失实作答样本。(3) 问卷内容。问卷内容为对于 30 个营销子系统信息采集品规的价格的判断，与模拟实验保持一致，采取五点式问题，同时考虑到部分品规卷烟会出现有商户未订购不了解的情况，增加了第六选项“我不卖不了解”。(4) 问卷发放计划。问卷依托问卷星平台进行发放，覆盖 754 户营销子系统信息采集商户，由客户经理负责转发并提供填写指导，发放前组织线上培训，明确问卷结构与答题要求，培训在微信群内完成。问卷连续投放 6 日，每日一轮来保证数据迭代连续性与反馈时效。

### 5.2 数据处理与问卷有效性检验

(1) 数据处理。收集到的问卷需要对数据进行清洗，去除无效、异常、不完整的专卖许可证号，随后利用数据清洗功能去除专卖证号重复填写的情况并保留最新一次提交时间的问卷，根据商户档位判断去除掉选择过多“我不卖、不了解”选项的问卷，将处理后的数据进行后续进一步的问卷有效性检验。

(2) 问卷有效性检验。模型预设商户有效回答率超过一半时可实现 99%精度修正，问卷设计中设置两类检验题以识别无效填答。其一为重复题一致性检验，如第 8 题与第 15 题针对黄金叶(硬红旗渠)的判断需保持一致；其二为固定答案题识别异常，如第 23 题设定黄金叶(硬帝豪)价格明显偏高，仅选择“低于 646.02”才视为有效。六轮问卷中检验题随轮次进行调整，用以剔除敷衍或认知偏差样本。

### 5.3 数据修正与最终结论

每日问卷数据经筛选后输入模型 V3.0，依据商户选择的价格判断项进行方向性修正。“过低”与“较低”触发向上调整，“较高”与“过高”引发下调，“持平”则不变。各项对应不同修正值，模型按品规分类逐项计算，逐步接近商户感知定价。

表 5-2 六轮问卷数据修正结果

卷烟品规	第一轮判断价格	第一轮修正价格 (第二轮判断价格)	第二轮修正价格 (第三轮判断价格)	第三轮修正价格 (第四轮判断价格)	第四轮修正价格 (第五轮判断价格)	第五轮修正价格 (第六轮判断价格)	第六轮修正价格
钻石(荷花)	350	340	337	335	334	335	335.43
中华(硬)	450	430	412	396	383	386	389.19
中华(双中支)	500	510	526	531	534	533	530.9
中华(软)	600	607	614	617	621	621	620.06

中华(金中支)	800	798	799	801	802	803	802.69
云烟(细支云龙)	150	149	149	148	148	147	147.51
云烟(软珍品)	220	217	213	209	207	208	209.32
玉溪(软)	220	216	213	210	208	209	209.97
苏烟(五星红杉树)	220	212	205	201	198	198	199.62
双喜(莲香)	120	127	133	138	142	141	140.77
南京(炫赫门)	180	187	193	199	203	203	200.74
利群(新版)	160	165	169	173	177	174	175.1
利群(蓝天)	180	183	186	189	191	191	191.01

第五轮的修正结果就是我们得到的最终营销子系统信息采集卷烟价格的99%精确度下的精准数据。

## 6 结尾

本课题聚焦卷烟价格数据集中的复杂性与精度问题，构建了以区间统计分析为核心的V3.0迭代修正模型，形成一套

系统化的数据采集与优化机制。以问卷调查为应用载体，面向754户信息采集客户开展六轮连续投放，利用模型修正反馈，完成五轮迭代与第六轮验证。模型实际运行结果表明，修正路径具备稳定性与收敛性，能有效提升价格数据的可信度与实时性，为行业数字治理提供有力支撑。

## 参考文献：

- [1] 李蕊.浅谈几个著名的大数定律及应用[J].科学咨询(科技管理),2010(12):64-65.
- [2] 滕泰,羿伟强,赵虹,等.全球大宗商品供求价格弹性分析[J].世界经济研究,2006,(06):59-64.
- [3] 许宏锋.高等教育管理审计.河海大学出版社,2007.11
- [4] 于秀娥主编.现代市场营销学.中国社会科学出版社,2010.01.
- [5] 张建英.博弈论的发展及其在现实中的应用[J].理论探索,2005,(02):36-37.